

# Vers la définition d'une méthode pour adapter les ressources terminologiques construites à partir de textes

Aurélien Picton<sup>1,2</sup> et Anne Condamines<sup>1</sup>

<sup>1</sup> CLLE-ERSS, CNRS & Université Toulouse 2

<sup>2</sup> Centre National d'Études Spatiales

[{prenom.nom}@univ-tlse2.fr](mailto:{prenom.nom}@univ-tlse2.fr)

## 1. Introduction

Les textes de spécialité peuvent être considérés comme des sources directes de la connaissance d'un domaine, notamment dans le cadre de la constitution de ressources terminologiques (glossaires, thesaurus, taxinomies, etc.). Or, les connaissances évoluent au cours du temps et les ressources deviennent vite inadéquates ou obsolètes. Il est donc nécessaire de pouvoir retracer cette évolution dans les textes pour les adapter. Cette problématique de l'évolution des connaissances prend un sens particulier dans le domaine spatial, dans lequel l'évolution se fait très rapidement et parfois à l'insu des protagonistes impliqués (ingénieurs, experts). À partir d'un besoin identifié par le Centre National d'Études Spatiales<sup>1</sup> et de l'observation du fait que

« (...) des textes scientifiques sur le même sujet produits à des moments éloignés dans le temps peuvent présenter des différences intéressantes non seulement au niveau de la conceptualisation de la discipline, mais aussi sur le plan de l'expression (...) » (Cabré, 1998, p141),

nous faisons l'hypothèse que des variations d'expression dans les textes (par exemple des changements de règles de formulations, de formes, de distribution, de fréquence, etc.) sont des indices de variations de sens/de concept dans le domaine traité et nous proposons de mettre en place une étude en diachronie à partir de textes du domaine pour appréhender cette évolution.

L'originalité de notre démarche réside dans deux éléments. Tout d'abord, compte-tenu de notre ancrage dans le domaine spatial où l'évolution des connaissances est très rapide, nous travaillons en diachronie « restreinte » : nos trois corpus couvrent une période de 8 ans, qui correspond à la date de parution de trois éditions d'un cours de technologie spatiale (cf. *infra*). Deuxièmement, notre principal objectif concerne l'aspect méthodologique. Nous visons en effet à mettre en place une méthode aussi généralisable que possible pour repérer les évolutions terminologiques dans les textes spécialisés.

Dans les sections suivantes, nous décrirons plus précisément la démarche générale mise en œuvre. Dans un premier temps, nous nous attarderons sur la nature du corpus pris en compte ainsi que certaines des contraintes sous-jacentes à sa constitution. Nous présenterons également les outils choisis pour accompagner la démarche de repérage des évolutions. Nous explorerons par la suite certaines observations de l'évolution des connaissances dans le corpus. Finalement, nous soulignerons deux constats centraux mis au jour à ce stade de notre recherche.

## 2. Cadre de l'étude

---

<sup>1</sup> Ce travail bénéficie d'un co-financement de thèse du Centre National de la Recherche Scientifique et du Centre National d'Études Spatiales.

Cette recherche part d'un besoin lié directement aux activités du Centre National d'Études Spatiales. En effet, il est des domaines où l'évolution des connaissances se fait de manière très marquée, où les évolutions scientifiques, les nouvelles découvertes viennent évidemment bouleverser, compléter ou affiner, l'état des connaissances. Le domaine spatial en est un très bon exemple, et ce, pour plusieurs raisons. Tout d'abord, le spatial, particulièrement en expansion depuis ces dernières décennies, est poussé par divers enjeux mondiaux et financiers qui le maintiennent dans une perpétuelle course en avant. Il s'agit d'un domaine de pointe où les technologies évoluent très rapidement. Deuxièmement, cette évolution rapide peut se produire à l'insu des protagonistes impliqués dans les projets spatiaux. Cet aspect est notamment probable dans le cas de projets de longue durée. En effet, il est courant de mettre en place des projets spatiaux échelonnés sur une dizaine ou une quinzaine d'années (par exemple dans le cas où on lance des sondes qui ne vont atteindre leur cible que des années plus tard). Entre le début et la fin du projet, les techniques évoluent « forcément » et ces changements peuvent se faire insidieusement. Cette situation peut poser différents types de problèmes : mauvaise communication entre ingénieurs et ingénieurs « juniors » qui arrivent en cours de projet, à un niveau de connaissances différent de celui des experts « seniors », oubli du contexte de connaissances dans lequel le projet a été initié, plus généralement, modification non-consciente du sens des termes. Cette situation a été repérée par le CNES comme potentiellement problématique. En tant que linguistes sollicités sur cette question, nous essayons de voir comment anticiper ces difficultés, tout d'abord en dressant un état des lieux des types d'évolutions qui peuvent apparaître.

Soulignons enfin que cette thématique de recherche permet d'aborder le lien entre textes et connaissances sous un nouvel angle. Nous inscrivons en effet notre recherche dans le cadre de la terminologie textuelle (Condamines, 2003) qui s'intéresse depuis longtemps à ce lien, mais d'un point de vue essentiellement synchronique (notamment dans le cadre de la constitution de ressources terminologiques à partir de textes). Ce qui est nouveau dans cette étude est que nous envisageons ce rapport textes/connaissance d'un point de vue diachronique (Aussenac-Gilles *et. al.*, à paraître). Nous pensons que les textes, de la même manière qu'ils « contiennent » des connaissances du domaine, contiennent également des indices, de nature diverse, pointant les évolutions de ces connaissances. Notre objectif est donc de mettre au jour une méthode qui permette de repérer et d'identifier dans les textes les évolutions de connaissances sur la base d'indices linguistiques.

### **3. Méthode d'analyse**

Comme nous l'avons présenté *supra*, l'hypothèse sous-jacente de ce travail est que les variations de sens/de concepts peuvent être mises au jour à partir du repérage de variations d'expression dans les textes. La première étape à mettre en œuvre est donc une étape d'élaboration d'un corpus qui réponde à un ensemble de contraintes liées à cet objectif.

#### **3.1. Constitution du corpus**

La difficulté (qui explique le soin que le chercheur, linguiste ou terminologue doit prendre à la réalisation de cette étape) réside dans la multiplicité des contraintes à prendre en compte pour obtenir un corpus exploitable dans cette optique de comparaison en terminologie diachronique. Nous en retenons principalement six :

- **contrainte de diachronicité « restreinte »** : compte tenu de l'évolution rapide du domaine et du besoin de repérage sur de courtes périodes (cf. 2.), nous devons disposer de textes rédigés à des dates différentes, éloignées de quelques années seulement (une quinzaine tout au plus) ;
- **contrainte d'homogénéité du genre** : il faut que les textes choisis soient représentatifs d'un même genre pour neutraliser l'influence que celui-ci pourrait avoir sur les variations repérables dans le corpus. À noter que cette contrainte est également pertinente en ce qui concerne le degré de spécialisation des textes (Bowker & Pearson, 2002, 28) ;
- **contrainte de variété des auteurs** : il faut éviter que les textes soient écrits par un seul et même auteur afin de circonscrire au maximum les variations particulières à un seul individu (idiosyncrasie) ;
- **contrainte de recours à des locuteurs natifs** : il faut privilégier les productions de rédacteurs natifs de la langue étudiée pour éviter que les phénomènes observables soient dus au manque de compétence langagière du rédacteur ;
- **contrainte de disponibilité des sources** : le choix des textes dépend étroitement des sources existantes, fiables, accessibles et non protégées par le secret industriel dans le cas de textes d'entreprise ;
- **contrainte de disponibilité d'un expert** : pour aiguiller le linguiste/terminologue dans sa tâche, la collaboration avec un expert du domaine est primordiale. Il faut donc privilégier l'intervention d'un expert qui puisse conseiller le chercheur dans sa tâche d'exploration. À noter que cette dernière contrainte ne va pas sans poser de nombreux problèmes d'appréhension et d'interrogation de la compétence d'un locuteur malgré la dimension temporelle et l'évolution de ses connaissances.

De manière générale, ces contraintes ne vont pas sans poser divers problèmes qui ne feront pas l'objet d'une description dans cet article, mais qui doivent rester présents à l'esprit du « diachronicien » lors de l'exploration de l'évolution.

Sur la base de ces différentes contraintes et avec l'aide du centre de documentation du CNES, nous avons construit un corpus de textes tirés du cours de Techniques et Technologies des Véhicules Spatiaux (TTVS). Ce cours est dispensé tous les ans par des experts du CNES à des semi-experts (Bowker & Pearson, 2002, 28) et les séminaires sont accompagnés de trois volumes de cours réédités tous les quatre ans depuis 1994 par le CNES, aux éditions Cépaduès. Ils sont rédigés au total par plus 80 experts spécialistes des domaines abordés. Le nombre de rédacteurs par chapitre peut varier de deux à plus d'une dizaine. Les trois volumes sont constitués d'une dizaine de chapitres englobant l'ensemble des domaines de spécialités du CNES : mécanique spatiale, localisation-navigation, télécommunication, etc. Parmi ces spécialités, notre choix s'est finalement porté sur les chapitres traitant d'optique et d'optoélectronique spatiales, notamment grâce à l'aide d'un expert en optique spatiale, rédacteur et relecteur de ce cours.

Au final nous disposons donc de trois sous-corpus correspondant à chaque édition du cours disponible :

1994	1998	2002
46 448 occurrences	78 656 occurrences	109 505 occurrences

Lorsque le linguiste travaille en diachronie sur un domaine de spécialité à partir de la comparaison de sous-corpus, plus ou moins éloignés dans le temps, il se trouve face à une alternative : soit il choisit de constituer un corpus « continu », c'est-à-dire un corpus constitué de textes le plus rapprochés dans le temps possible, par exemple un corpus de textes journalistiques qui

contiendrait l'ensemble des faits divers édités chaque jour pendant dix ans. Ainsi, il fait l'hypothèse que son corpus contiendra l'ensemble des phénomènes d'évolution de la langue sur une période donnée. Soit il décide de travailler sur un corpus que nous qualifierons de corpus « à intervalles », où il dispose alors de plusieurs textes édités à des périodes différentes entre lesquelles aucune ressource représentative d'un état de langue intermédiaire aux époques sélectionnées n'est disponible.

Dans le second cas, se pose le problème du découpage dans le temps, c'est-à-dire de la partition du corpus : sur quelles bases choisir les états de langue à observer ? Quel découpage adopter entre différentes périodes ? Généralement, celui-ci peut se faire de manière arbitraire (le chercheur suppose que le laps de temps choisi est suffisant pour observer une évolution, (Møller, 1998)) ou sur la base d'un événement extralinguistique (le linguiste étudie les variations avant et après un événement historique tels que les attentats du 11 septembre 2001, (Drouin *et. al*, 2006)).

Dans un contexte d'entreprise tel que le nôtre, il est difficile de parvenir à réunir des sources continues telles que des dépêches quotidiennes. Nous avons donc fait le choix dans cette étude de travailler à partir d'un corpus « à intervalles » : le corpus contient trois sous-corpus édités tous les quatre ans. L'originalité de notre corpus est ainsi que les intervalles de temps qui existent entre les sous-corpus n'ont pas été choisis arbitrairement ou sur la base d'un événement extérieur, mais ils émanent directement des besoins du domaine et de son rythme d'évolution, comme l'illustre cette citation de D. Assémat dans la préface de l'édition 2002 du TTVS (TTVS, 2002) :

« Voici l'édition 2002 du cours de Techniques et Technologies des Véhicules Spatiaux. Le cycle de 4 ans se trouve ainsi institutionnalisé (94 - 98...) et le lien avec le cycle de vie de nos développements technologiques n'est pas fortuit. (...) En quatre ans, des changements profonds de stratégie de développement des segments spatiaux sont intervenus. ».

Enfin, nous terminerons cette description de notre corpus d'étude par une dernière remarque : bien que les intérêts du CNES ont été ciblés sur la base de l'évolution des connaissances dans le cadre des projets spatiaux de longue durée, notre corpus n'est pas, à ce stade de notre recherche, un « corpus projet » (c'est-à-dire un corpus constitué de textes liés à des projets, tels que des rapports de spécification par exemple). En effet, dans la mesure où le TTVS est construit selon une certaine visée didactique, nous espérons que dans un premier temps, ce genre de texte nous permettra de relever des indices d'évolution plus difficiles à déceler dans un corpus hautement technique écrit par des experts pour d'autres experts.

### **3.2. Outillage**

La démarche que nous mettons en place pour l'observation des indices d'évolution dans les textes est une démarche linguistique outillée.

Tout d'abord, les termes sont extraits grâce à l'analyseur syntaxique Syntex développé par D. Bourigault (Bourigault *et. al.*, 2005). Les termes extraits (noms et syntagmes nominaux, adjectifs et syntagmes adjectivaux, verbes et syntagmes verbaux) sont filtrés automatiquement à l'aide d'un programme Perl. Ce programme permet d'enlever une partie du bruit de l'analyse Syntex : candidats-termes trop longs, mauvais découpage du syntagme, etc. (par exemple, termes tels que *fréquence de modulation plus élevée, coût très élevé des instruments spatiaux, and Space Environment*, etc.). Ce premier tri automatique est suivi par un second tri, manuel cette fois-ci, mené par nos soins et ponctuellement par un expert en optique. Ce second tri permet d'éliminer le

bruit restant que le programme Perl ne permet pas de repérer (par exemple, *grand*, *qualité*, etc. qui ne sont pas des termes spécifiques au domaine de l'optique).

Une fois triés, les termes sont ensuite visualisés dans une base de données prototype nommée TermOnto (Bourigault *et. al*, 2004). Cette base de données facilite la navigation : visualisation des termes en listes, en contextes, comparaison de fréquences entre sous-corpus, etc. La base de données offre également la possibilité d'interroger la liste de termes par le biais de requêtes (regroupement de termes de préfixe commun, regroupement de termes de fréquence croissante, etc.) et permet ainsi une exploration paramétrable et facilitée. L'analyse des termes en contextes est également assistée par le concordancier MonoConc Pro, développé par M. Barlow.

À ce stade, il est bien évident que la démarche outillée que nous mettons en place ne se veut pas une démarche de repérage automatique. Il s'agit d'une assistance automatique pour un dépouillement manuel des indices. La question de l'évolution dans les textes est en effet une question récente en terminologie textuelle, les techniques ne sont pas assez matures et la question de l'évolution des connaissances pose un ensemble de problèmes non résolus qui ne permettent pas d'envisager du « tout automatique » (par exemple, à quel moment peut-on considérer qu'un concept a assez évolué pour parler d'un nouveau concept ?). D'ailleurs, à notre connaissance, une seule tentative d'automatisation du repérage de l'évolution a été mise en place par A. Tartier (Tartier, 2004) et les résultats semblent assez décevants.

#### **4. Premières observations**

Lorsque nous avons présenté notre hypothèse de travail, nous avons mentionné le fait que des variations d'expression dans les textes pouvaient être le signe de variations de sens/concept. Ce que nous entendons par « variations d'expression » renvoie aussi bien à des différences de fréquences, de formes, de distribution (Condamines et al., 2004) qu'à l'apparition de contextes évolutifs (contextes qui marquent explicitement une évolution, une nouveauté, une amélioration, etc.). Autrement dit, seront considérées toutes les variations linguistiques repérables dans les textes constituant des indices potentiels à mettre au jour.

Nous présentons dans les sections qui suivent quelques uns des premiers indices révélés par notre étude.

##### **4.1. Présence de contextes évolutifs**

Il est possible de dégager dans les textes des « indices langagiers de nouveauté pour détecter ce qui change dans le contenu des textes eux-mêmes » (Ibékwé-Sanjuan, 2005). Plus généralement, sont présents dans les textes différents indices textuels auxquels on peut intuitivement associer une idée d'évolution. Il peut s'agir d'adverbiaux (*actuellement*, *récemment*, *depuis {date}*, etc.), d'adjectifs (*nouveaux*, *prometteur*, *ancien*, etc.), de noms (*prototype*, *innovation*, *développement*, etc.), de prépositions (*au stade de*, etc.), de verbes (*abandonner*, *remplacer*, etc.), de comparatifs (*plus/moins {Adj} que*, etc.), etc. D'un point de vue sémantique, ces indices portent aussi bien une valeur de nouveauté (*nouveau*) que d'obsolescence (*ancien*) ou de succession/remplacement (*remplacer par*). Notons également que la particularité de ces indices est qu'ils s'appliquent à des éléments de niveaux différents : lexical (adjectif, noms) ou discursif (adverbe).

Une fois choisis, ces indices sont projetés dans les textes où on les observe pour ajuster notre sélection. Certains indices sont ajoutés après plusieurs itérations et certains sont enlevés dans la mesure où ils apportent trop de bruit (par exemple, si *gain* a été sélectionné dans un premier temps, nous avons choisi de l'éliminer au vu du nombre de contextes non pertinents qu'il fait émerger : *Ce*

*processus ne peut toutefois pas être trop poussé, car le mécanisme d'amplification (**gain**) de la photodiode est d'autant plus bruyant que ce **gain** est élevé.*

Le point commun de ces indices est donc qu'ils peuvent indiquer l'arrivée d'une technique, les tendances, les concepts susceptibles de « prendre le devant de la scène » dans le domaine, etc. Il est donc intéressant de projeter ces contextes dans les textes pour observer les termes qui y gravitent.

Pour ce faire, nous considérons comme fenêtre d'observation le paragraphe qui permet d'observer ces indices, quel que soit leur niveau de portée, discursif ou lexical. Pour illustrer, prenons un exemple de contexte évolutif pertinent :

*« Signalons qu'**actuellement** de **nouveaux** produits, photodiodes avalanches à très grandes surfaces et détecteurs hybrides entre PM et photodiode avalanche, **commencent à apparaître** : **meilleurs gains** (10 puissance 3), **grandes surfaces** (plusieurs dizaines de mm<sup>2</sup>) et des facteurs de bruit réduits (1,1 à 1,2). » (TTVS 2002).*

Ce paragraphe est très riche en indices évolutifs (notés en gras) et à la lecture, il est permis de penser que les photodiodes avalanches à très grandes surfaces et les détecteurs hybrides sont des technologies nouvelles qui seront les technologies courantes de demain. Ceci doit néanmoins être discuté avec un expert dans la mesure du possible.

Signalons également que plus un paragraphe est riche en marqueurs d'évolution, plus il est pertinent. En effet si plus de 90% des contextes qui contiennent au moins 4 marqueurs sont pertinents, ce taux descend à 22% dans le cas de contextes ne contenant qu'un seul marqueur.

Les avantages du repérage de ce type d'indices d'évolution est que celle-ci est simple à mettre en œuvre et qu'elle permet de pointer effectivement vers des évolutions potentielles. Cependant, le choix des indices reste subjectif et leur portée dépend très certainement du genre du texte. Ainsi, la dimension didactique des textes étudiés explique sans doute la nécessité pour les rédacteurs de contextualiser leur propos en inscrivant les concepts dans l'histoire de la discipline. De la même manière, ces indices seront certainement moins présents dans un corpus de textes techniques tels que des rapports de spécification où la dimension didactique est totalement absente

## **4.2. Évolution morphologique**

Un autre type d'indices observables pour repérer une évolution potentielle est l'évolution morphologique.

Comme le souligne D. Assémat dans la préface du cours de 2002 :

*« La miniaturisation est la règle. Le CNES a créé des filières de satellites (mini et micro) pour disposer de produits sur étagère, pas chers, capables d'assurer les missions souhaitées. Nos orientations techniques et technologiques se font autour de la réduction physique des éléments permettant de réaliser des fonctions encore plus performantes et, à moindres coûts. »*

Nous retrouvons donc logiquement l'émergence de nombreux dérivés et composés formés à partir des préfixes micro- et mini- : *mini-satellite, mini-TES, microbolomètres, micro-lasers*, etc.

D'un point de vue compositionnel, il est également fréquent de rencontrer des termes composés nouveaux, issus d'un terme existant spécifié par un adjectif : *MOS (Metal Oxyde Semiconductor)* donne *CMOS (Complementary MOS)*, qui lui-même donne *BiCMOS (Bipolar*

CMOS). L'interprétation qui peut en découler (et cette interprétation doit être encore une fois validée par des experts du domaine) est que CMOS et BiCMOS sont les dernières nées des technologies MOS, le BiCMOS étant lui-même une amélioration des technologies CMOS.

D'un point de vue flexionnel, il est possible d'observer l'apparition de marques du pluriel sur certains sigles (et les sigles offrent dans ce cas un champ d'observation non négligeable ; en effet, ce type d'évolution en morphologie flexionnelle ne peut apparaître que sur certains types d'unités, tels que les sigles ou les noms propres). Bien que cette tendance à marquer graphiquement le pluriel des sigles soit typiquement anglo-saxonne (et s'observe d'ailleurs essentiellement sur des sigles d'origine anglo-saxonne), elle n'en reste pas moins un indice révélateur de la maturation et de la bonne intégration d'une technique ou d'une technologie. Par exemple, si *ASIC* (*Application Specific Integrated Circuit*) est un sigle invariable dans le sous-corpus de 1994, on observe dès 1998 des occurrences telles que *des ASICs*, marquant un meilleur ancrage de ce sigle dans le système de la langue.

### 4.3. Évolution graphique

Il apparaît à plusieurs reprises dans notre corpus une tendance à la siglaison, ou du moins à la réduction de termes à certaines lettres qui le composent. Prenons l'exemple du *rapport signal sur bruit*.

On remarque que cinq variantes coexistent dans notre corpus pour désigner un même concept : le rapport signal sur bruit (« Rapport établi entre la puissance du signal d'origine et la puissance du bruit, à un point donné du système. » (www.granddictionnaire.com)). Dans le tableau suivant, nous répertorions ces cinq variantes et indiquons pour chacune leur proportion d'apparition par rapport à l'ensemble des variantes possibles, et ce, pour chacun des sous-corpus. Notons que cette proportion d'apparition est directement inspirée du coefficient d'implantation proposé par J. Quirion pour une mesure de l'implantation terminologique des termes (Quirion, 2003). Pour chaque colonne, la variante préférée pour nommer ce concept est mise en évidence en gras.

	1994	1998	2002
<b>rapport signal à bruit</b>	<b>50</b>	21,4	19
<b>rapport signal sur bruit</b>	10	14,3	16,7
<b>rapport signal/bruit</b>	5	4,8	2,4
<b>signal/bruit</b>	0	2,4	2,4
<b>rapport S/B</b>	10	21,4	26,2
<b>S/B</b>	25	<b>35,7</b>	<b>33,3</b>

On s'aperçoit qu'en 1994, le terme *rapport signal à bruit* domine nettement (il est choisi dans 50% des cas). Mais dès 1998, date de la deuxième édition, la répartition des variantes est moins homogène et la variante dominante est une forme réduite de *rapport signal à bruit* : *S/B* (concurrencée de près par une variante légèrement plus longue : *rapport S/B*). Cette tendance semble se confirmer en 2002 où la proportion d'occurrences de *rapport signal à bruit* diminue encore un peu (soit une baisse 11,21% entre 1998 et 2002).

Dans le cas de l'évolution graphique, il ne s'agit pas vraiment de l'évolution d'un concept, mais de l'évolution de la forme du concept qui peut, à moyen ou long terme, provoquer une perte de connaissance, d'autant plus si la forme « développée » venait à disparaître, opacifiant un peu plus l'identification du concept désigné.

#### 4.4. Constats

La mise en avant de ces premiers indices d'évolution potentielle entraîne au moins deux constats à prendre en compte dans la démarche de repérage.

Tout d'abord, nous devons souligner que, dans la plupart des cas, chacun de ces indices pris séparément peut être discutable quant à son rôle dans le repérage d'une évolution.

Prenons l'exemple de BiCMOS, que nous récapitulons dans le tableau suivant :

1				2				3
Paradigme : concept spécifié				Stabilisation graphique				Présence de contextes évolutifs
	94	98	02		94	98	02	
MOS (Metal Oxide Semiconductor)	17	18	20	bi-CMOS	3	0	0	[1994] En ce qui concerne la technologie <u>bi-CMOS</u> , ce n'est qu'en 1993 que nous venons d'avoir accès aux logiciels [...] de conceptions de circuits. [2002] Aujourd'hui, [...], ce sont les technologies <u>CMOS</u> qui font l'objet d'améliorations incessantes. [...] On constate que les technologies <u>BiCMOS</u> ne sont qu'une amélioration des technologies <u>CMOS</u> existantes.
CMOS (Complementary + MOS)	6	6	7					
BiCMOS/bi-CMOS (Bipolar + CMOS)	3	4	4					

La première colonne (1) illustre un aspect mentionné *supra* : la spécification du concept. L'ajout d'un modifieur à un terme existant entraîne généralement la création d'un concept spécifié (marqué par un lien d'hyperonymie). Il est donc important de pouvoir retracer ses séries morphologiques. Cependant, cette série seule ne permet pas de postuler une évolution des connaissances entre 1994, date du premier sous-corpus, et 2002, date du plus récent. En effet, bien qu'il y ait vraisemblablement spécification d'un concept, si cette spécification a eu lieu plusieurs années auparavant, elle n'est pas pertinente par rapport à notre objectif. Différentes techniques affiliées peuvent en effet coexister plusieurs années au sein d'un domaine. La spécification met en évidence dans ce cas la technique la plus récente en non pas une technique nouvelle.

La deuxième colonne (2) indique que *BiCMOS* a subi une modification graphique entre 1994 et 1998 : le tiret inscrit à l'origine disparaît et la graphie de ce terme se stabilise. Ce type de phénomène est un indice potentiel d'évolution. Mais encore une fois, à lui seul, il n'est pas suffisant (cette stabilisation pourrait être due à une correction orthographique lors de la relecture du cours, par rapport à la première édition ; il ne faut en effet pas oublier qu'une seconde édition est également toujours une amélioration de la précédente).

La troisième colonne (3) montre quant à elle la présence de contextes dits évolutifs. Dans ces extraits, les technologies BiCMOS sont explicitement présentées comme émergentes en 1994 et



comme étant des améliorations des technologies CMOS en 2002. Mais encore une fois, bien que ces contextes pointent explicitement des évolutions, ils n'en sont pas moins des indices d'évolution discutables. En effet, dans certains cas, les arguments des rédacteurs du cours ne font pas forcément l'unanimité : si certains présentent une technique particulière comme une clef pour l'avenir du domaine, d'autres peuvent ne pas s'accorder sur ce point. Il faudrait aussi sans doute distinguer les indices qui évoquent une évolution à venir, toujours hypothétique (*prometteur...*), de ceux qui rendent compte d'une évolution avérée (*X a été abandonné,...*).

Au vu de ces éléments, un aspect à souligner est le fait que les évolutions de connaissances sont marquées dans les textes par différents biais. Il semble donc qu'une évolution est, du moins dans certains cas, révélée par un ensemble de variations et non pas une seule. Dans la mesure du possible, il faut donc privilégier des « faisceaux d'indices » pour postuler une évolution potentielle dans le domaine. Malgré tout, la validation finale de l'évolution doit prendre en compte l'avis d'un expert.

Un second constat, qui découle des différents types d'indices présentés est que les évolutions de connaissances repérables dans les textes n'interviennent pas toutes à un même degré de conscience. En effet, dans des cas tels que : « *une nouvelle classe d'instruments devient nécessaire : l'interférométrie* », la nouveauté de l'interférométrie est une évolution consciente. Les experts ont totalement conscience du fait que cette classe d'instruments vient d'apparaître dans leur domaine de spécialité. À l'inverse, pour reprendre l'exemple du terme *rapport signal à bruit*, l'évolution graphique, révélatrice de l'importance de cette notion dans le domaine, est un processus beaucoup moins conscient et la perte de connaissances associée à la réduction est d'autant plus risquée. Le glissement graphique, et l'opacité qui en découle, est une évolution très insidieuse. Il en est de même, selon nous, pour les glissements de sens. Et c'est justement là un argument de plus au besoin de repérage de ce type de changement.

## 5. Perspectives

Dans cet article, nous avons essayé de montrer les premiers résultats obtenus dans le cadre de notre travail de thèse. Nous avons présenté certains indices linguistiques d'évolutions que nous avons pu mettre au jour à partir d'une démarche de repérage outillée sur des textes du domaine de l'optique spatiale. À ce stade, nous ne proposons pas de conclusion à proprement parler, mais plus d'une série de perspectives à mettre en œuvre pour approfondir notre recherche.

Tout d'abord, comme nous l'avons souligné dans la dernière partie, un type d'évolution particulièrement intéressant mais difficile à repérer, concerne tout type d'évolution inconsciente. Pour essayer de la mettre au jour, nous pensons que l'étude de la distribution des termes pourra apporter des pistes intéressantes (Condamines *et. al.*, 2004). L'hypothèse posée ici est qu'un terme employé dans des contextes différents à des périodes différentes aura subi une modification au cours du temps. Pour le moment, cette étape n'est que très peu avancée et pose de nombreuses questions méthodologiques. Nous avons cependant pu déjà repérer quelques cas intéressants tels que l'évolution du terme *éléments parasites*. Bien entendu, les éléments parasites ne sont pas un concept nouveau dans le domaine de l'optique qui doit les gérer depuis sa création. L'intérêt ici est que l'étude de la distribution de ce terme permet de dégager un changement de point de vue sur ce concept au cours du temps. En effet, le seul contexte d'apparition de ce terme en 1994 est *extraire* ~. Or, en 2002, les contextes se diversifient : ~ *limitent*, *falloir maîtriser* ~, *risquer d'augmenter* ~, ~ *êtres dus à*, et montrent une prise de conscience et une instance plus forte sur le besoin et les moyens mis en œuvre pour pallier l'effet indésirable des parasites. Dans ce cas, mais cela reste une

perspective à explorer, il est permis de penser qu'il y a bien eu une évolution dans les moyens disponibles ou envisagés pour leur maîtrise. Et ce changement de distribution peut révéler un bouleversement plus important qu'il ne peut y paraître dans le domaine.

Une deuxième perspective est de mieux exploiter le parallélisme de nos sous-corpus. En effet, dans la mesure où il s'agit de trois éditions successives d'un même cours, ces sous-corpus semblent idéalement comparables. En alignant certaines parties sur le modèle des corpus comparables, il sera peut-être possible de faire ressortir des différences pertinentes.

Enfin, à terme, nous souhaitons mettre en œuvre les différents indices repérés sur un corpus associé à un projet spatial de longue durée puisque c'est pour ce type de domaine que l'évolution est apparue comme potentiellement problématique.

## 6. Références

- Aussenac-Gilles, N., Condamines, A. & Sèdes, F. (éds) (à paraître) *Textes et ressources terminologiques et/ou ontologiques : évolution et maintenance*, Revue I3 (Information, Interaction et Intelligence).
- Bourigault D., Aussenac-Gilles N. & Charlet J. (2004) « Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas », In *Revue d'Intelligence Artificielle (RIA), Techniques Informatiques et structuration de terminologiques*, Pierrel J.-M. et Slodzian M. (Ed.), Paris, Hermès, 18(1), pp87-110.
- Bourigault D., Fabre C., Frérot C., Jacques M.-P. & Ozdowska S. (2005), « Syntex, analyseur syntaxique de corpus », In *Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelle, atelier EASy : campagne d'évaluation des analyseurs syntaxiques*, Dourdan, France, 6-10 juin 2005, II, pp17-20.
- Bowker, L., & Pearson, J. (2002) *Working with specialized language : a practical guide to using corpora*, Routledge, London/New York.
- Cabré, M. T. (1998) *La terminologie : théories, méthodes et applications*, Armand Colin, Les Presses de l'Université d'Ottawa, Ottawa.
- Condamines, A. (2003) « Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques », Mémoire d'Habilitation à Diriger les Recherches, In *Carnet de Grammaire*, Rapports Internes de l'ERSS (Équipe de Recherche en Syntaxe et Sémantique), 13.
- Condamines, A., Rebeyrolle, J. & A. Soubeille (2004) « Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus », In *Actes d'Euralex'2004*, Lorient, France, 6-10 juillet 2004, pp547-557.
- CNES (1994) *Cours de Techniques et Technologies des Véhicules Spatiaux*, Centre National d'Études Spatiales, Toulouse, 3 volumes.
- CNES (1998) *Cours de Techniques et Technologies des Véhicules Spatiaux*, Éditions Cépaduès, Toulouse, 3 volumes.
- CNES (2002) *Cours de Technologies Spatiales : Techniques et Technologies des Véhicules Spatiaux*, Éditions Cépaduès, Toulouse, 3 volumes.
- Drouin, P., Paquin, A. & N. Ménard (2006) « Extraction semi-automatique des néologismes dans la terminologie du terrorisme », In *Actes des 8èmes Journées internationales d'Analyse statistique des Données textuelles (JADT-2006)*, Besançon, France, 19-21 avril 2006, pp389-400.
- Ibékwe-Sanjuan, F. (2005) « Repérage et annotation de nouveautés dans les écrits scientifiques », In *Actes du colloque « Indices, Index et Indexation »*, Lille, France, 3-4 novembre 2005.

- Quirion, J. (2003) *La mesure de l'implantation terminologique : proposition d'un protocole. Étude terminométrique du domaine des transports au Québec*, In *Langues et sociétés*, 40, Montréal, Office Québécois de la Langue Française.
- Tartier, A. (2004) *Analyse automatique de l'évolution terminologique : variantes et distances*, Thèse de doctorat en informatique, Université de Nantes.